Article

https://doi.org/10.1038/s41467-024-55285-9

Inter-chromosomal transcription hubs shape the 3D genome architecture of African trypanosomes

Received: 10 July 2024

Accepted: 6 December 2024

Published online: 23 December 2024

Check for updates

Claudia Rabuffo ^{1,2}, Markus R. Schmidt ^{1,2}, Prateek Yadav ^{1,2}, Pin Tong³, Roberta Carloni^{3,4}, Anna Barcons-Simon ^{1,2}, Raúl O. Cosentino^{1,2}, Stefan Krebs ⁵, Keith R. Matthews ⁴, Robin C. Allshire ³ & T. Nicolai Siegel ^{1,2}

The eukaryotic nucleus exhibits a highly organized 3D genome architecture, with RNA transcription and processing confined to specific nuclear structures. While intra-chromosomal interactions, such as promoter-enhancer dynamics, are well-studied, the role of inter-chromosomal interactions remains poorly understood. Investigating these interactions in mammalian cells is challenging due to large genome sizes and the need for deep sequencing. Additionally, transcription-dependent 3D topologies in mixed cell populations further complicate analyses. To address these challenges, we used high-resolution DNA-DNA contact mapping (Micro-C) in Trypanosoma brucei, a parasite with continuous RNA polymerase II (RNAPII) transcription and polycistronic transcription units (PTUs). With approximately 300 transcription start sites (TSSs), this genome organization simplifies data interpretation. To minimize scaffolding artifacts, we also generated a highly contiguous phased genome assembly using ultra-long sequencing reads. Our Micro-C analysis revealed an intricate 3D genome organization. While the T. brucei genome displays features resembling chromosome territories, its chromosomes are arranged around polymerase-specific transcription hubs. RNAPI-transcribed genes cluster, as expected from their localization to the nucleolus. However, we also found that RNAPII TSSs form distinct inter-chromosomal transcription hubs with other RNAPII TSSs. These findings highlight the evolutionary significance of inter-chromosomal transcription hubs and provide new insights into genome organization in T. brucei.

To fit inside the nucleus of a cell, the long polymers of DNA that contain a cell's genetic information must be tightly packaged¹. Research over the past few decades has shown that DNA is not randomly packaged and that the 3D genome organization inside the

nucleus regulates vital cellular processes, including gene expression and DNA repair^{2–4}. Chromosome conformation capture (3C)-based techniques, which measure the frequency of DNA-DNA contacts between any two DNA loci, have revealed a large number of distinct

¹Division of Experimental Parasitology, Faculty of Veterinary Medicine, Ludwig-Maximilians-Universität München, 82152 Planegg-Martinsried, Germany. ²Biomedical Center Munich, Division of Physiological Chemistry, Ludwig-Maximilians-Universität München, 82152 Planegg-Martinsried, Germany. ³Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom. ⁴Institute of Immunology and Infection Research, University of Edinburgh Edinburgh. ⁵Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany. ⁶e-mail: n.siegel@lmu.de

structures, collectively referred to as 'contact domains'. Although it has long been recognized that contact domains vary in size and function, the terminology used to describe different contact domains has changed over time. To reduce ambiguity, in our study we follow the definitions proposed by Beagan and Phillips-Cremins⁵. Here, 'chromatin domains' are defined as any triangle of enhanced contact frequency tiling the diagonal of a DNA-DNA contact matrix.

At the largest scale, eukaryotic DNA is typically organized into active and inactive regions that segregate into A and B compartments⁶. 'Compartment domains' are thus defined as chromatin domains whose boundaries coincide with inflection points in the A/B compartmentalization signals, and compartment boundaries are identified by computing the eigenvectors of the interaction matrices^{7,8}.

At the intermediate scale, genomes can be organized into 'topologically associating domains' (TADs). Beagan and Phillips-Cremins define TADs as contact domains whose formation is driven by cohesinmediated loop extrusion^{5,9,10}. These domains are flanked by CCCTCbinding factor (CTCF), a protein that restricts the movement of cohesin and thereby defines the boundaries of TADs¹¹⁻¹³. TAD boundaries can be identified by calculating an insulation score that reflects the amount of interaction between two TADs and other contact domains.

At the finest scale, high-resolution 3C-based approaches in metazoans have revealed so-called 'chromatin loops', defined as a point of enriched contacts in Hi-C or Micro-C heat maps, which appear as a dot (a series of adjacent pixels with enhanced contact frequency relative to the local chromatin domain structure). Such loops can represent distinct interactions between enhancers and promoters that may be important for gene regulation¹⁴⁻¹⁷.

Additionally, microscopy^{18,19} and methods capable of capturing long-distance and multivalent interactions, such as split-pool recognition of interactions by tag extension (SPRITE)²⁰, have shown that specific inter-chromosomal interactions can organize around distinct nuclear bodies. For example, RNA polymerase I (RNAPI)-transcribed rDNA genes, encoded on several different chromosomes, localize within the nucleolus²¹. Although dispersed in the linear genome, tRNA genes transcribed by RNA polymerase III (RNAPIII) can form spatial clusters²²⁻²⁴, which facilitate the coordination of transcription and processing of tRNAs, crucial for protein synthesis. Similarly, RNA polymerase II (RNAPII)-transcribed genes have been found to cluster in structures, sometimes referred to as transcription factories^{25,26}, condensates²⁷, pockets²⁸ or hubs²⁹, which are discrete nuclear sites where active transcription occurs. These clusters can bring together multiple genes from different regions of the genome to be transcribed simultaneously.

The organization of genes into inter-chromosomal hubs has also been reported to be important beyond coordinating transcription and to contribute to efficient RNA processing. Trypanosoma brucei is a highly divergent, unicellular parasite responsible for sleeping sickness in humans and nagana in cattle in sub-Saharan Africa. Microscopy and Hi-C data from T. brucei have suggested that the 3D nuclear organization is important for RNA maturation³⁰⁻³². In those studies it was found that highly expressed T. brucei genes, such as the active VSG gene and tubulin genes, make frequent inter-chromosomal interactions with a locus that produces the noncoding spliced leader RNA that is trans-spliced to the 5⁻ end of every mRNA and is thus crucial for mRNA maturation³³. These findings led to a model in which a distinct 3D nuclear organization may enhance the maturation of specific transcripts. Similarly, in mammalian cells, it was observed that nuclear speckles represent major structural hubs that organize interchromosomal contacts. Such structures reduce the spatial distance of genes from these speckles, thereby increasing RNA splicing rates³⁴.

What remains unclear is whether these specific interchromosomal interactions identified by microscopy, SPRITE, and 3Cbased methods represent special cases or broader principles of global

genome into distinct inter-chromosomal hubs, e.g., around functionally related DNA elements, may prove especially important in organisms lacking some of the well-studied contact domains such as TADs and loops between promoters and enhancers (E-P loops).

genome organization. It seems plausible that the organization of the

However, most high-resolution 3C-based analyses have been performed in organisms with distinct intra-chromosomal organization, such as mammals^{35,36}, *Drosophila*³⁷, and yeasts³⁸⁻⁴⁰, making it difficult to judge the evolutionary importance of inter-chromosomal interactions. In addition, transcription-dependent topologies, such as clustering of genes transcribed by RNAPII, may be transient features difficult to detect in data generated from mixed populations, given that nuclear DNA organization is highly dynamic in mammals⁴¹⁻⁴⁴. Furthermore, the large size of the human, mouse, or Drosophila genome make the generation of high-resolution contact maps costly. As a result, most studies do not have the necessary depth to investigate genome-wide DNA contacts at high resolution or require specialized strategies to focus on a subset of genomic regions^{45,46}.

Here, to investigate the role of inter-chromosomal hubs in the absence of TADs and E-P loops, we generated high-resolution Micro-C contact maps in T. brucei. Unlike more complex eukaryotes, T. brucei lacks enhancers and CTCF⁴⁷, and a previous Hi-C study points to the absence of TAD-like structures⁴⁸. In addition, unusually for a eukaryote, its ~9000 genes transcribed by RNAPII are organized into a few hundred polycistronic transcription units (PTUs). Moreover, transcription initiation of the PTUs is unregulated, removing the interference of dynamic transcription regulation in the interpretation of architectural data^{49,50}. These features, in addition to its relatively small genome of 35 Mb, make T. brucei an ideal organism to use highresolution 3C methods to investigate the evolutionary role of interchromosomal hub structures and to identify transcription-dependent nuclear topologies. A challenge in analyzing 3D genome organization in less well-studied organisms has been a lack of chromosome-scale genome assemblies, which complicates the analysis of 3C data and often leads to erroneous data interpretation. Particularly, extended gaps, collapsed repeat regions, and incorrectly scaffolded contigs can lead to misinterpretations of 3C data. Despite significant improvements in the quality of the *T. brucei* genome assembly over the last 10 years^{48,51,52}, the current assembly of the Lister 427 genome (version 10) contains 103 gaps and 74 collapsed repeats. Thus, for this study we generated ultra-long nanopore reads at ~350x coverage, enabling us to greatly reduce the number of gaps and collapsed repeats in the new genome assembly. Next, we established high-resolution Micro-C assays, allowing us to determine the parasite's genome architecture at unprecedented resolution. This revealed an intricate 3D genome organization around polymerase-class specific transcription hubs.

Our observations indicate that not only do RNAPI-transcribed rRNA genes interact with each other, but RNAPII TSSs also come together to form inter-chromosomal transcription hubs. Similarly, we find that RNAPIII-transcribed tRNA genes interact with other tRNA genes, albeit in a more selective manner than RNAPII TSSs.

Results

Ultra-long nanopore reads resolve most gaps and collapsed repeats in the T. brucei genome assemblies

To investigate the presence of polymerase class-specific inter-chromosomal hub structures, we sought to determine whether regions transcribed by RNAPI, RNAPII, or RNAPIII interact more frequently with themselves than with other regions of the genome. In T. brucei, 45 of the 120 annotated RNAPI-transcribed rRNA genes are located in repetitive regions that are collapsed or absent in the genome assembly (Lister 427 version 10). We therefore aimed to fill gaps and correct collapsed regions to their true size using ultra-long Oxford Nanopore Technology (ONT) sequencing reads.

Approximately ~350-fold coverage of the T. brucei Lister 427 strain genome was generated to correct the previous fully phased Lister 427 assembly (version 10)⁵¹. Previous data⁵¹ were merged with six new sequencing runs, with read N50s ranging from ~40 kb to ~440 kb (Supplementary Table 1). The merged data allowed the partial or complete closure of 85 of the 103 gaps, ranging in size from 2 to 50 kb (Fig. 1a, b). To assess the completeness of gap closure, we used our long ONT reads and extracted the first and last 1.5 kb of each read to generate virtual paired reads (VPRs) (Fig. 1c, d and Supplementary Fig. 1a). Since the distance between the paired reads is known from the ONT read length, the distance between genome-aligned VPRs can be used to evaluate whether a gap has been completely closed and whether a collapsed repeat region has been extended to the correct length. Using the VPR analysis, we were able to validate the complete closure of 58 gaps. The remaining gap closures resulted in partial closures or could not be verified by VPRs. Similarly, we used the VPRs to identify 74 regions where assembly size was inconsistent with the size of the ONT reads. Of these 74 regions, we corrected 63 to their correct length, shortening 13 by up to 12 kb and extending the remaining 50 by up to 17 kb (Fig. 1a, b).

In order to accurately map our sequencing data, it is critical to take into account the unusual linear structure of the trypanosome genome. *T. brucei* has a diploid genome with individual chromosomes containing both homozygous core regions and subtelomeric regions with a very high level of heterozygosity. When mapping sequencing data, it is important to represent the highly similar regions only once to avoid mapping sequencing reads to multiple loci. Therefore, we generated a working version of our new genome containing both haplotypes (A and B) of the highly heterozygous subtelomeric regions assembled as separate contigs, but only haplotype A of the diploid core regions.

Importantly for the subsequent analyses, in the new genome assembly, all TSSs, rDNA loci and 95% of the tRNA genes are in correctly assembled regions (i.e., more than 1kb from an open gap or collapsed region). In addition, 16 of the 21 annotated centromeres are now fully assembled, ranging in size from 9 to 55 kb. These improvements to the *T. brucei* Lister 427 genome assembly significantly enhance its continuity and make it well-suited for genome wide sequencing-based analyses, including 3C-based studies.

Micro-C reveals organization of RNAPII transcription start sites into inter-chromosomal hubs

Previous Hi-C analysis, in which DNA was crosslinked with formaldehyde and fragmented by a specific restriction enzyme, revealed important features of the *T. brucei* genome organization within the nucleus. For example, while no evidence for TADs was found, interchromosomal interactions were detected that appear to enhance trans-splicing of highly expressed genes^{30,32,48}. However, the resolution of Hi-C is limited by the cutting frequency of the restriction enzyme used. To detect DNA-DNA interactions at the highest possible resolution and to evaluate the presence of interactions between polymerasespecific transcription start sites, we established Micro-C^{35,39} in *T. brucei*. Using an asynchronous culture with 80-90% of cells in interphase, DNA is fragmented with micrococcal nuclease (MNase), which cuts DNA between nucleosomes, allowing a resolution of -150 bp. Additionally, a second crosslinking step was added to allow a better capture of longdistance interactions and reduce the signal-to-noise ratio⁷.

After sequencing, to correctly adjust for the discrepancy between the physical number of genomic regions and the number of contigs in the genome assembly due to the unusual structure of the *T. brucei* genome, we corrected our Micro-C data for ploidy by duplicating interactions from underrepresented subtelomeric regions (for details see Methods and ref. 53).

Visual inspection of the Micro-C data revealed a striking pattern of enriched interactions between RNAPII TSSs, suggesting the presence

of distinct chromatin loops and RNAPII hub structures (Fig. 2a). To systematically identify loops, i.e., enriched interactions between distant bins on the same chromosome, we used Mustache⁵⁴, a tool for multi-scale detection of chromatin loops from Hi-C or Micro-C data. Using Mustache, we identified 367 loops. Of the 367 loops identified, 269 connected two sites (bins) annotated as TSS. Analysis of the remaining 98 loops revealed that they either connected sites with low read coverage, resulting in high noise, or connected a TSS to a bin that was immediately adjacent to a TSS. Taken together, this analysis suggests that it is predominantly regions with RNAPII TSSs that are involved in loop formation. To confirm this observation, we performed an aggregate analysis, averaging interactions between pairs of RNAPII TSSs located on the same chromosome, and again found that these sites interact much more frequently than control regions (regions of identical size located 20 kb downstream on the same chromosome) (Fig. 2b). Of the 1914 TSS-TSS interactions pairs for which interaction frequencies could be determined, 64.8% had $a \ge 1.25$ -fold higher interaction frequency than their corresponding control region.

Classical chromatin loops and most promoter-enhancer interactions occur within chromosomes. However, when we extended our analysis to include pairs of TSSs located on different chromosomes, we still detected enriched interactions between TSSs (Fig. 2c). Of the 14,412 TSS-TSS inter-chromosomal interactions pairs for which interaction frequencies could be determined, 76.8% had a \geq 1.25-fold higher interaction frequency than their corresponding control region. Although inter-chromosomal TSS-TSS interactions occurred at a lower frequency than interactions between TSSs located on the same chromosome, these results suggest the formation of inter-chromosomal hub structures where RNAPII transcription start regions cluster (Fig. 2c).

Unlike TSSs in yeast and more complex eukaryotes, TSSs in *T. brucei* are not marked by well-defined nucleosome-depleted regions⁵⁵. Nonetheless, our data were normalized by Iterative Correction⁵⁶ (IC) to ensure that the observed TSS-TSS interactions did not result from an artifact related to MNase digestion bias. Additionally, we re-analyzed previously generated Hi-C data, where DNA was digested using the 4-base cutter MboI. While individual TSSs with enriched contacts were only clearly identifiable in our Micro-C data (Fig. 2a), aggregate analyses around all sites of enriched contacts based on Micro-C data (n = 367) also showed enrichment in the Hi-C data (Fig. 2d). The detection of these interactions by Hi-C and Micro-C indicates that the observed interactions between TSSs are not due to biases in MNase digestion.

RNAPII TSSs segregate into distinct compartment domains

The best-studied mechanism for chromatin loop formation in metazoans involves the cooperation of CTCF, a DNA-binding protein, and cohesin, a ring-shaped protein complex, in a process called loop extrusion¹⁰. In this process, cohesin complexes are thought to actively extrude loops of chromatin through their ring-like structure until they encounter boundary elements such as CTCF, where extrusion is blocked. However, non-metazoans including *T. brucei* lack CTCF⁴⁷. Furthermore, in contrast to studies in other non-metazoans such as budding and fission yeasts³⁸⁻⁴⁰ or in closely related species such as *T. cruzt*^{77,58}, studies on the nuclear architecture in *T. brucei* have not identified any TAD-like structures⁴⁸.

However, to better understand the nature of the observed interchromosomal RNAPII TSS hub structures, we used the new Micro-C data and performed a systematic analysis to identify compartment domains and other contact domains based on changes in eigenvector value and insulation score, respectively.

Eigenvectors calculated from the correlation matrix at 10 kb resolution using FAN-C⁸ suggested non-exclusive compartmentalization at three levels. As described previously⁴⁸, we found that the transcribed chromosome core regions and the untranscribed subtelomeric



Fig. 1 | **A highly contiguous** *T. brucei* **genome assembly was generated with ultra-long nanopore reads. a** An overview of the features and regions on the megabase chromosomes of the Lister 427 *T. brucei* assembly that have been modified. Allele 'A' and 'B' are shown on the top and bottom of each chromosome, respectively. Modified regions (gaps or collapsed regions, fully or partially corrected) are marked in blue. Regions with open gaps or collapsed repeats that could not be corrected and remained unmodified are marked in red. **b** Sizes of closed

gaps (top) and change in size of corrected repeat regions (bottom). **c** Illustration of our strategy to confirm the correctness of the modifications made to the genome. Modified regions are considered corrected if there are at least 2 VPRs with less than 5 kb difference between the distance where the VPRs map on the assembly and their distance based on ONT read. **d** Correctness evaluation of modified regions in the new genome assembly (version 12).



Fig. 2 | Micro-C reveals interactions among RNAPII transcription start sites.
a IC-normalized and ploidy-corrected heat maps of Hi-C (bottom left) and Micro-C (top right) of megabase chromosomes 1-11 (haplotype A) at 50 kb resolution (left) with each chromosome displayed with the core region in white, subtelomeric regions in light gray and centromeres displayed as black lines; part of chromosome 10 of the same heat maps is displayed at 5 kb resolution (right), with polycistronic transcription units of chromosome 10 illustrated as black arrows (top).
b Aggregate analysis of Micro-C data around pairs of intra-chromosomal TSSs

(n = 2342 TSS-TSS pairs); control loci are obtained from the same pairs of loci shifted downstream by 20 kb. **c** Aggregate analysis of Micro-C data around pairs of inter-chromosomal TSSs (n = 20,008 TSS-TSS pairs); control loci are obtained from the same pairs of loci shifted downstream by 20 kb. **d** Aggregate peak analysis of Hi-C and Micro-C data around chromatin loops (i.e., pairs of bins with high interaction frequency) identified in the Micro-C matrix using Mustache (with a *p*-value threshold of 0.02 and a sparsity threshold of 0.85; n = 367).

regions fold into distinct compartments. In addition, the eigenvector values suggest that in some cases the two chromosome arms separated by the centromere may fold into distinct compartments. Third, our data suggest that RNAPII TSSs form a compartment distinct from the rest of the core regions (Fig. 3a and Supplementary Fig. 6).

Next, we used FAN-C⁸ to calculate the insulation score, a measure of the average frequency of interactions that extend beyond a given bin (here 10 kb). Typically, a low insulation score marks the boundaries of contact domains, including TADs. However, low insulation scores can also be caused by analysis artifacts, such as regions of low mappability caused by repeat regions, gaps in the genome assembly, or incorrectly scaffolded contigs. Our data suggest that low insulation scores correlate with low read mappability in all but one region (Fig. 3b and Supplementary Fig. 7). Interestingly, this region on chromosome 8 is flanked by two sites that represent boundaries between PTUs, contain tRNA genes, and are enriched in cohesin (Fig. 3b), similarly to what has been found in closely-related species⁵⁸.

Thus, our Micro-C data suggest that while the *T. brucei* genome appears to be organized into distinct compartmental domains, with almost all regions of low insulation coinciding with sites of low mappability, our data did not provide evidence for the presence of contact domains separated by low insulation scores, such as TADs and TAD-like structures, except in one case.



Fig. 3 | **Transcription start sites localize to distinct kilobase-scale compartments. a** Examples of correlation matrices for chromosomes 4, 5 and 11 (haplotype A) and corresponding second eigenvector calculated from the Micro-C data analyzed with the distiller pipeline at 10 kb resolution (using FAN-C⁸). Each chromosome is displayed under the eigenvector plot with the core region in grey and subtelomeric regions in dark red; centromeres are represented as black dots.

b Examples of domain boundaries identified manually with insulation scores calculated from the Micro-C data analyzed with the distiller pipeline at 2.5, 5 and 10 kb resolution using FAN-C⁸ (top); mappability of the region, TAD-like domains identified with HiCExplorer¹⁰⁶, no visibility regions (i.e., regions where signal was removed by IC normalization, for details see Methods), tRNA genes, SCC1 ChIP-seq fold enrichment and PTUs are displayed (bottom).

Interactions among RNAPII TSSs are independent of cohesinmediated loop extrusion

The clustering of RNAPII TSSs into distinct compartments, the lack of evidence for the existence of TAD-like structures in *T. brucei*, and the relatively high frequency of inter-chromosomal interactions between TSSs argue against cohesin-mediated loop extrusion as the main driver in establishing the observed TSS hub structures.

However, since cohesin is often found to be enriched at TSSs and cohesin-mediated chromatin loops are frequently anchored at TSSs⁵⁹, we asked whether cohesin contributes to the formation of the observed TSS hubs. Depletion of cohesin followed by Micro-C assays would be the most direct way to determine if it contributes to hub formation, but depletion of the major cohesin subunit SCC1 by RNAi leads to rapid cell death in *T. brucei*^{60,61}. Instead, the distribution of cohesin across the *T. brucei* genome was reassessed using existing SCC1 ChIP-seq data⁴⁸ and our improved genome assembly. Our reanalysis revealed that SCC1 is enriched at transcription termination sites (TTSs) but depleted at TSSs (Fig. 4). Because PTUs are found on either strand of the genome, their TTSs can be convergent (cTTS) or, when PTUs are oriented head-to-tail, they can be located upstream of

adjacent start sites as single TTS (sTTS). We found cohesin to be enriched at both convergent and single TTSs. In fact, cohesin levels gradually increase along PTUs, with the highest cohesin levels at the TTSs (Fig. 5a), as if its distribution is dependent on elongating RNAPII. In contrast, it is the TSS at the beginning of the PTU that engage in interactions with other TSSs (Fig. 5a). These observations suggest that a role of cohesin in the establishment of the observed TSS hubs is unlikely.

The fact that such interactions were not observed for RNAPII TTSs or for the body of PTUs (Figs. 4 and 5b) suggests that the specialized chromatin structure associated with RNAPII TSSs may be important for the formation of these RNAPII TSS hubs. Unlike in most eukaryotes, in *T. brucei* RNAPII TSSs are not marked by well-positioned nucleosomes and narrow peaks of H2A.Z containing +1 nucleosomes⁵⁵. In contrast, *T. brucei* RNAPII TSSs exhibit high levels of the histone variants H2A.Z and H2B.V, histone acetylation at positions H4K2, H4K5, and H4K10, and tri-methylation at position H3K4. These histone variants and modifications are enriched across regions of -10 kb in width, encompassing the first few genes of every PTU^{62–66}. We re-analyzed previously generated H2A.Z MNase-ChIP-seq data⁵⁵ using our improved genome



Fig. 4 | **Cohesin is depleted from sites engaging in TSS-TSS interactions. a** Top panel: schematics of polycistronic transcription unit types (head-to-head and head-to-tail) with divergent and single transcription start sites (dTSS and sTSS, in green) and single and convergent transcription termination sites (sTTS and cTTS, in red).

Bottom panels: for each of the sites, aggregate Micro-C signal (observed/expected), metaplot of SCC1 ChIP-seq⁴⁸ fold enrichment, RNAPII ChIP-seq⁶⁶ fold enrichment and H2A.Z MNase-ChIP-seq⁵⁵ enrichment.

assembly. This analysis revealed that a much higher percentage of Micro-C contact anchor sites overlapped with peaks of H2A.Z MNase-ChIP-seq peaks (69.8%) than with SCC1 ChIP-seq peaks (1.4%; Fig. 5c). This finding suggests that it may be the unique chromatin structure found at TSSs that drives clustering of TSSs rather than cohesin-mediated loop extrusion.

RNAPI transcribed genes cluster together selectively

RNAPII transcription in *T. brucei* is unregulated and constitutively active^{49,50}. In contrast, RNAPI-transcribed genes coding for the major surface antigens, the variant surface glycoproteins (VSGs), are highly

regulated. More than 2000 VSG-encoding genes are present in the *T. brucei* genome, but only one of these VSG genes is transcribed at any given time³². Moreover, FISH and Hi-C analyses indicate that the majority of transcriptionally repressed VSG genes cluster into several distinct loci^{48,67}. In contrast, the single actively transcribed VSG gene is located away from these clusters of transcriptionally inactive VSG genes³⁰.

Perhaps relevant to the silencing of RNAPI-transcribed VSG genes is the fact that not all rRNA genes are actively transcribed simultaneously in other organisms and that their activation can be selective. Some rRNA genes are more transcriptionally active than others in





Fig. 5 | **Cohesin levels gradually increase along PTUs. a** Metaplot of SCC1 ChIPseq⁴⁶ fold-enrichment (yellow) and of RNAPII ChIP- seq⁴⁶ fold enrichment (purple) along polycistronic transcription units (n = 208). **b** Aggregated plot of interaction frequencies calculated from Micro-C data between pairs of PTUs (n = 208). PTUs were scaled to the same length and extended by 10% up- and downstream. **c** UpSet

plot showing the chromatin loops (i.e., pairs of bins with high interaction frequency) identified in the Micro-C matrix at 5 kb resolution (n = 367), where both loci of a pair overlap with annotated TSSs (n = 298), H2A.Z MNase-ChIP-seq⁵⁵ peaks (n = 592; see Methods), annotated TTSs (n = 309), SCC1 ChIP- seq⁴⁸ peaks (n = 381; see Methods), and/or annotated tRNA genes (n = 137).

chromosome conformation and nuclear localization data²²⁻²⁴ suggest

that these genes associate in 3D space within nuclei. Such clustering of

tRNA genes has been hypothesized to increase the efficiency of their

transcription by concentrating the necessary transcription factors and

RNAPIII machinery in specific nuclear regions⁷⁷. T. brucei RNAPIII

transcribes tRNA genes, snRNAs located in close proximity to tRNA

genes, and 5S rRNA genes78-81. To identify any prominent spatial con-

tacts between RNAPIII-transcribed genes, we performed a virtual 4C

analysis using every cluster of tRNA genes as separate viewpoints.

tRNA clusters were defined as groups of adjacent tRNA genes where

each gene is separated from the next by less than 10 kb. In contrast to

our analysis of RNAPII- and RNAPI-transcribed genes, we found that

while approximately 25% of the tRNA regions contacted one or more

RNAPIII-transcribed regions, the others did not appear to engage in

DNA-DNA contacts with other RNAPIII-transcribed loci (Fig. 6b, bot-

mammalian cells, and these active rRNA genes tend to cluster together within the nucleolus⁶⁸⁻⁷¹. Similarly, in *T. brucei*, promoters of different rRNA genes show differential activity^{72,73}.

To determine whether interactions between *T. brucei* rRNA genes occur across the genome, we plotted our Micro-C data in a 4C-like manner using RNAPI-transcribed rDNA regions as the selected viewpoints. Since rRNA genes are often arranged in tandem arrays along chromosomes, we defined rDNA regions as groups of adjacent rDNA loci separated by less than 10 kb. This resulted in the assignment of 16 rDNA regions distributed across the 22 megabase-sized chromosomes. Significance was assigned to virtual 4C peaks using a two-sample Kolmogorov–Smirnov test (KS-test; see Methods). However, since virtual 4C peak sizes are influenced by factors such as the size of the viewpoint and the mappability of viewpoint and target regions, we also examined interactions qualitatively.

Our analysis revealed a network of interactions in which rDNA loci interact preferentially in specific groups. Specifically, the rDNA loci on chromosomes 1, 3, and 7 form one cluster, while two rDNA genes on the subtelomeric region of chromosome 11 A interact with the rRNA clusters on chromosomes 9 and 10 (Fig. 6a). Similar to the polymerase class-specific clustering we saw for RNAPII TSSs, we observed that the RNAPIII-transcribed 5S rRNA genes located on chromosome 8 interact only minimally with RNAPI-transcribed rRNA loci, and we suggest that these interactions are a consequence of the spatial clustering of rRNA genes in the nucleolus. The repetitive nature of rRNA gene arrays prevented us from conclusively assessing patterns of selective transcription. Thus, it remains to be seen whether the selective clustering of rDNA into different hubs correlates with the transcriptional activity of these loci.

Micro-C reveals pairwise interactions between RNAPIIItranscribed tRNA genes

In most eukaryotes, including *T. brucei*, tRNA genes are scattered throughout the genome and are often found in arrays on multiple chromosomes^{74–76}. Despite this dispersed linear organization, previous

th rDNA locitom track, and Supplementary Fig. 8).DNA loci onMicro-C suggests that centromeres form multiple clustersA genes onMicro-C suggests that centromeres form multiple clustersth the rRNAGiven the unusual organization of the genome around polymerasepolymeraseclass-specific hubs, we decided to extend our analysis to study theorganization of other important genomic features often found tointeractinteract in specific clusters, such as centromeres.To characterize centromere-centromere interactions in moredetail, we improved the assembly of centromeric sequences andannotated the centromeres using information from the kinetochoreprotein KKT2 ChIP-seq data⁸² or the presence of centromere-specificCIR147 repeats^{83,84}. Using these data, we were able to annotate the

CIR147 repeats^{83,84}. Using these data, we were able to annotate the centromeres of 21 of the 22 megabase-sized chromosome homologs. No KKT2 signal was detected on chromosome 10 A. Our annotations indicate that centromere regions vary in length from 3 to 50 kb and are composed of different classes of repeat elements, as suggested based on topoisomerase II cleavage⁸⁵ and confirmed by our consensus repeat sequence analysis (Table 1 and Supplementary Table 3). Because the repetitive nature of centromeric DNA makes it difficult to reliably map reads to specific centromeric loci, we generated virtual 4C plots for



Fig. 6 | **rRNA and tRNA gene loci form polymerase class-specific clusters. a** Top: Virtual 4C interaction profiles based on Micro-C data (10 kb resolution) showing the interaction frequencies of rDNA regions (defined as groups of adjacent rDNA loci separated by less than 10 kb) used as viewpoints with the Tb427 genome version 12, haplotype A. Significant interactions as calculated using a two-sample KS-test were marked below each track (pink: p < 0.05; red: p < 0.01). Bottom: rDNA regions are highlighted in green; red boxes represent centromeres; each chromosome is displayed with the core region in white and subtelomeric regions in light gray. **b** Top: Virtual 4C interaction profiles based on Micro-C data (5 kb resolution) showing the interaction frequencies of tRNA regions (defined as groups of adjacent tRNA loci separated by less than 10 kb) used as viewpoint with the Tb427 genome version 12, haplotype A. Significant interactions as calculated using a two-sample KS-test were marked below each track (pink: p < 0.05; red: p < 0.01). Bottom: tRNA regions are highlighted in yellow; red boxes represent centromeres; each chromosome is displayed with the core region in white and subtelomeric regions in light gray.

Table 1 | Size, location, period size and class of centromeres in the newly assembled genome Lister 427 version 12

Chromosome	Size (bp)	Contig type	Period size (bp) and class
1A	43,790	core	146, class I
1B	50,856	core	146, class I
2A	14,757	core	29, class II
2B	8800	core	29, class II
3A	24,693	core	120, class IV
3B	24,343	core	120, class IV
4A	33,158	core	148, class III
4B	54,643	core	148, class III
5A	>21,486*	core	147, class III
5B	20,864	core	147, class III
6A	>3745*	core	58/136, class I
6B	>3029*	core	58/136, class I
7A	39,869	core	30, class II
7B	39,480	core	30, class II
8A	19,090	core	147, class III
8B	18,625	core	147, class III
9A	>2741*	3' subtelo	39, class I
9B	50,482	3' subtelo	39/78, class I
10B	18,287	3' subtelo	49, class I
11A	>3806*	3' subtelo	49, class I
11B	43,917	3' subtelo	20/59, class I

Centromeres that are not fully assembled are marked with*.

each centromere (Fig. 7a) and repeated the analysis using 5 kb regions immediately upstream or downstream of the bins containing the centromere repeats as viewpoints (Supplementary Fig. 9 and 10). We reasoned that interactions of flanking regions with a specific centromere in a distance-dependent manner would strongly suggest that these centromeres interact. Using this approach, we found that centromeres 1, 2, 3, and 7 clustered together (Fig. 7a) and that centromere 4 clustered with centromere 5 (Fig. 7a and Supplementary Figs. 9, 10). Moreover, we used fluorescence in situ hybridization (FISH) with probes hybridizing to the CIR147 repeats at centromeres 4, 5 and 8 to quantify the number of clusters in the nucleus of T. brucei. In the absence of centromere clustering, we would expect to detect as many foci by FISH as there are physical copies in the genome. In the case of CIR147 repeats, we would therefore expect 6 foci in G1 cells and 12 in dividing cells. Instead, we detected <4 foci in the majority of cells (91%; Fig. 7b). These results suggest that centromeres 4, 5 and 8 are indeed colocalizing. As control we used FISH to visualize the region transcribed into spliced leader RNA, a single repeat region located on both homologs of chromosome 9. Here, we find the number of FISH loci to match the number of genomic sites, indicating that these two sites do not cluster (Supplementary Fig. 11). Taken together, these findings suggest that not only transcribed regions but also centromeres cluster into specific hubs (Fig. 7c).

Discussion

Inter-chromosomal interactions play an important role in chromatin domain formation, gene regulation and genome stability⁸⁶. However, previous studies have predominantly focused on intra-chromosomal structures and their impact on transcriptional regulation. Given the unique organization of trypanosome genes into a few hundred transcription units and their constitutive transcriptional activity^{49,50}, this unicellular parasite represents an excellent model to explore the principles of inter-chromosomal interactions.

To study the nuclear architecture of *T. brucei* in detail, we generated a highly contiguous genome assembly, closing most gaps and

correcting 85% of the collapsed repeats in the 11 megabase-sized chromosomes. We then performed a high-resolution Micro-C analysis of bloodstream *T. brucei*, generating >170 million uniquely mapping reads. Given the *T. brucei* genome size of 35 Mb and the human genome size of 3.3 Gb, the resolution of our study is comparable to a study with -1.6 trillion reads of the human genome when analysing inter-chromosomal interactions.

On a global scale, based on eigenvector values, our data suggest a compartmentalization of the *T. brucei* genome at three levels: 1) The transcribed chromosome core regions and the untranscribed subtelomeric regions fold into distinct compartments, as we had previously observed⁴⁸. 2) In some cases, the centromeres appear to act as compartment boundaries, with chromosome arms folding into distinct compartments. 3) RNAPII TSSs form a compartment that is distinct from the rest of the core regions. Given that the entire core regions of chromosomes are transcribed in *T. brucei*, we did not expect the core regions to segregate into A and B compartments as described in other organisms. However, we were surprised to find that RNAPII TSSs appear to segregate into a separate compartment, and we propose that this may be a result of the RNAPII TSSs clustering.

To detect contact domains other than compartments, we used our Micro-C data to calculate insulation scores. As in a previous study⁵⁷, we detected numerous sites with low insulation scores, marking putative contact domain boundaries. However, we also found that most of these low insulation score sites coincided with sites of low read mappability, implying that the interaction frequencies calculated for these sites are not reliable. Thus, while the output of the contact domain callers should be treated with caution, as it may simply be an artifact caused by low mappability regions, some of the many low mappability sites may contain repetitive regions that represent bona fide contact domain boundaries. One way to address this issue in the future would be to perform Pore-C analyses⁸⁷. The longer reads generated in such studies should address the issue of low unique mappability. In summary, while we think it is possible that future studies provide evidence for the existence of contact domain boundaries, we see no evidence that these domains are generated by loop extrusion, a feature of TAD-like structures.

The high resolution of our dataset revealed both conserved and unusual polymerase class-specific inter-chromosomal hubs. While interactions among RNAPI-transcribed genes within the nucleolus appear to be ubiquitously conserved, and interactions among RNAPIII-transcribed tRNA genes have been described in many organisms^{22–24,58}, we also observed an extensive pattern of inter-chromosomal RNAPII TSS interactions. Our observations are consistent with previous immunofluorescence analyses of RNAPII⁸⁸ and H4K10ac, a histone mark enriched at RNAPII TSSs⁶², which revealed a punctate nuclear pattern suggesting clustering. In other organisms, interactions among RNAPII-transcribed genes have been observed mainly for specific gene subsets associated with increased splicing in nuclear speckles^{20,30,89}.

Taken together, our Micro-C data suggest that intra-chromosomal interactions leading to organization into compartments enclosing the transcribed core regions or the transcriptionally repressed subtelomeric regions provide the global genome organization. In contrast, inter-chromosomal interactions provide a much finer additional level of interactions between specific loci, such as RNAPII TSSs, rRNA genes, or between the single actively transcribed VSG gene and the SL array, that may be important for RNA transcription and processing.

Although our data revealed the presence of intra- and interchromosomal clustering of RNAPII TSSs, because we measured pairwise rather than multivalent interactions, the true size of these hubs and their variability between cells remains unknown. For example, although our data suggest that most TSSs interact with multiple other TSSs, it is still unclear whether these interactions occur simultaneously or involve different loci in different cells. Furthermore, the driving factors behind these inter-chromosomal hubs remain unknown. One



Fig. 7 | *T. brucei* genome form widespread inter-chromosomal interactions with centromere clustering. a Top: Virtual 4C interaction profiles based on Micro-C data (10 kb resolution) showing the interaction frequencies of a subset of centromeres used as viewpoints (black arrowheads and green background) with the Tb427 genome version 12, haplotype B. Significant interactions as calculated using a two-sample KS-test were marked below each track (pink: p < 0.05; red: p < 0.01). Bottom: Each chromosome is displayed with the core region in white and

subtelomeric regions in light gray. Centromeres are marked by red rectangles; rDNA loci are marked in green. **b** Fluorescence in situ hybridization (FISH)-based quantification of CIR147 repeats (located on chromosomes 4, 5 and 8) clusters (N=1; analyzed nuclei: 171). Cells displaying no signal were excluded. **c** Model of *T*. *brucei* genome organization. *T. brucei* chromosomes make frequent centromerecentromere interactions, interactions between RNAPII TSSs, the active VSG gene and the splice leader locus, rRNA genes and some tRNA genes (not displayed). hypothesis is that the accumulation of RNA polymerases and/or nascent RNAs at TSSs may facilitate the formation of transcriptional condensates, as observed in other species²⁵⁻²⁹. Alternatively, the distinct profile of histone variants and modifications at TSSs may influence compartmentalization^{62,65,66,90}. We have previously shown that reduced H2A.Z deposition at TSSs does not affect *T. brucei* RNAPII PTU transcript levels, but instead shifts transcription initiation to more accessible sites in chromatin⁶⁵. Whether such a shift affects the level of interaction between TSSs remains to be seen.

Despite the extensive network of inter-chromosomal interactions, the high frequency of intra-chromosomal interactions observed in this study suggests some degree of chromosomal territoriality. Because our data were generated from asynchronous cells, our results are consistent with a scenario in which chromosomes occupy distinct territories during mitosis and subsequently undergo global rearrangements⁹¹. Another possibility is that chromosomes move within the nucleus, frequently interacting with intra-chromosomal regions while dynamically establishing contacts with other chromosomes.

In this study, we focused on bloodstream form parasites, the stage that infects the mammalian host. Although Hi-C data exist for insect-stage parasites³⁰, no systematic comparison of genome architecture has been performed for this stage. Thus, it remains to be seen how the genome organization changes during the life cycle of the parasite. Furthermore, in this and previous studies, we have focused on the diploid set of the 11 megabase-sized chromosomes. However, *T. brucei* also contains 5-6 intermediate chromosomes and -100 minichromosomes. Because the intermediate and minichromosomes are not yet fully assembled, they have been excluded from genome architecture analyses remains to be determined in future studies.

Our results highlight the importance of studying genome organization in evolutionarily divergent organisms, particularly those with small genomes where deep sequencing is affordable, and reveal an intricate pattern of inter-chromosomal DNA-DNA interactions.

Methods

Trypanosoma brucei cell culture and growth

In this study, two cell lines were used: *T. brucei* wildtype (WT) Lister 427 bloodstream form; and a double-selection *T. brucei* cell lines derived from the Lister 427 bloodstream-form MITat 1.2 isolate⁹². Both cell lines were grown in HMI-11 medium at 37 °C and 5% CO₂. In the latter cell line, a neomycin resistance gene in bloodstream-form expression site (BES) 17 and a puromycin resistance gene in BES 1 allowed the selection for a homogenous cell population that expressed either *VSG-2* from BES 1 or *VSG-13* from BES 17. This cell line was grown in HMI-11 supplemented with either 20 µg/mL neomycin (also referred to as N50 cells) or 1 µg/mL puromycin (referred to as P10 cells).

gDNA extraction for ONT sequencing

For nanopore sequencing, six sequencing libraries were obtained from P10 cells or *T. brucei* wildtype Lister 427 cells. For each library, gDNA extraction was performed using High Molecular Weight DNA Extraction Kit from Cells and Blood (NEB #T3050). The samples were sequenced using a GridION X5 or PromethION 24 platform on a FLO-PRO114M flowcell using MINKNOW software. Sequencing reports and details of each sequencing libraries are shown in Supplementary Table 1.

Closing gaps and expanding repeats in the assembly

First, an assembly error in BES 2 of the previously fully phased version 10 of the *T. brucei* Lister 427 strain genome assembly⁵¹ was corrected and the contigs names were shortened to generate the version 11 of the assembly (Tb427v11). In order to close gaps and expand collapsed repeats in the Tb427v11 assembly, SAMBA⁹³ was run with the ONT

reads generated in this study (Supplementary Table 1) and the ONT reads from Girasol et al.⁹⁴. It was confirmed that small contigs are not oriented the wrong way around, an error commonly found in assemblies scaffolded using Hi-C data (Supplementary Fig. 1a-c)⁹³. Further, duplicated sequences next to gaps were removed (Supplementary Fig. 2a-d), collapsed repeats were converted to gaps that were successively replaced with the expanded sequences using SAMBA (Supplementary Fig. 3), and gaps in homozygous regions were closed by running SAMBA separately on alleles A and B. 'samtools stats' was used to compute the error rates for each newly generated sequence and all sequences from the Tb427v11 assembly (Supplementary Fig. 4). The step-by-step approach is detailed in Supplementary Note 1.

The newly assembled regions were annotated using Companion⁹⁵. To transfer the annotations from unmodified regions of the Tb427v11 assembly to the improved one, the original annotations' coordinates were transformed based on the cumulative size difference of the modified regions. To avoid read-alignment ambiguities arising from both alleles of a homozygous region being present in an assembly, a working version of the assembly was generated using only one allele (allele A) of each homozygous core region and both alleles of all heterozygous subtelomeric regions, defined as the version 12 of the assembly (Tb427v12).

Centromere annotation and consensus repeat sequence analysis

KKT2 ChIP-seq data⁸² were mapped to the Tb427 version 12 (Tb427v12) genome build using Bowtie2⁹⁶ (version 2.4.2), with duplicate reads removed using SAMtools⁹⁷. KKT2 enrichment was identified using MACS2⁹⁸ (version 2.2.7.1) broad peak call. Each main chromosome exhibited a large KKT2 enrichment cluster, ranging from approximately 3 kb to 50 kb. The enrichment patterns observed between both biological replicates were used to determine the centromere locations. Consensus sequences on each *T. brucei* chromosome were determined using Tandem Repeats Finder⁹⁹.

Micro-C

The Micro-C protocol was adapted from Krietenstein et al.³⁵. We generated 3 biological replicates and 2-3 technical replicates for each biological replicate for a total of 8 replicates. Cells in log phase (0.6- 1×10^{6} cells/mL) were collected and resuspended in 12.5 mL of 1X trypanosome dilution buffer (TDB; 0.005 M KCl, 0.08 M NaCl, 0.001 M MgSO4, 0.02 M Na2HPO4, 0.002 M NaH2PO4, 0.02 M glucose) per 100 million cells. Cells were fixed with 1% formaldehyde for 20 min at room temperature by adding 1.25 mL of freshly made formaldehyde solution (50 mM HEPES-KOH (pH 7.5), 100 mM NaCl, 1 mM EDTA (pH 8.0), 0.5 mM EGTA, 11% formaldehyde). The crosslinking reaction was quenched by adding 2.5 M Tris buffer (pH 7.5) to a final concentration of 0.75 M at room temperature for 5 min, then on ice for 15 min. Cells were washed twice with 1X TDB and then crosslinked again by addition of freshly made DSG solution to the final concentration of 3 mM for 45 min at room temperature. The crosslinking reaction was quenched by addition of 2.5 M Tris buffer as described above. Cells were washed once in 1X TDB and snap-frozen in liquid nitrogen in aliquots of 50 million cells.

Frozen cells were resuspended in 200 μ L of cold 1X TDB per 50 million cells. After 20 min incubation on ice, cells were collected by centrifugation (10,000 g, 5 min at 4 °C), washed once with 400 μ L and once with 200 μ L complete MB#1 buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 5 mM MgCl2, 1 mM CaCl2, 0.2% NP-40, 1x Roche cOmplete EDTA-free (Roche diagnostics, 04693132001)) and resuspended in 50 μ L of complete MB#1 buffer.

Chromatin was fragmented for 10 min at 37 °C with MNase concentrations pre-titrated to yield mostly mononucleosomal fragments (Merck, #N3755). The digestion was stopped by addition of 4 mM EGTA and incubation at 65 °C for 10 min. Here, chromatin aliquots were pooled for further processing. 200 million cell-equivalent per replicate yielded the best results. Chromatin was collected by centrifugation (16.000 g. 10 min at 4 °C), washed with 500 uL of 1x NEBuffer 2.1 (NEB, #B7202S), collected by centrifugation (16,000 g, 10 min at 4 °C), and resuspended in 45 µL of NEBuffer 2.1. DNA ends were dephosphorylated by addition of 5 µL of rSAP (NEB, #M0203) and incubation at 37 °C for 45 min, then the reaction was stopped by incubation at 65 °C for 5 min. DNA overhangs were generated by adding 40 µL of "overhang mix" (5 µL of 10x NEBuffer 2, 0.2 µL of 200X BSA, 2 µL of 100 mM ATP (Thermo Fisher, #R0441), 3 µL of 100 mM DTT, 29.8 µL of H2O, 8 µL of Large Klenow Fragment (NEB, #M0210L), 2 µL of T4 PNK (NEB, #M0201L)) and incubation at 37 °C for 15 min. The DNA overhangs were filled by adding 100 µL of "fill-in" mix (25 µL of 0.4 mM Biotin-dATP (Invitrogen, #19524-016), 25 µL of 0.4 mM BiotindCTP (Invitrogen, #19518-018), 1 µL of 10 mM dGTP, 1 µL of 10 mM dTTP, 10 µL of 10X T4 DNA Ligase Reaction Buffer (NEB #B0202S), 0.5 µL of 200X BSA (NEB, #B9000S), 37.5 µL of H2O) and incubation at 23 °C for 3.5 h with interval shake. The reaction was stopped by addition of 12 µL of 0.5 M EDTA and incubation at 65 °C for 20 min.

Chromatin was collected by centrifugation (16,000 g, 10 min at 4 °C), washed with 1 mL of 1X T4 DNA Ligase Reaction Buffer (NEB #B0202S) and collected by centrifugation (16,000 g, 10 min at 4 °C). The chromatin pellet was resuspended in 1.8 mL of "proximity ligation" mix (1X T4 DNA Ligase Reaction Buffer (NEB #B0202S), 1X BSA (NEB, #B9000S), 11,200 U of NEB T4 Ligase (NEB, #M0202L)) and incubated rotating at room temperature for 2.5 h.

After proximity ligation, the chromatin was collected (>18,000 g, 10 min at 4 °C), resuspended in 200 μ L of 1X of NEBuffer 1 (NEB, #B7001S) and 200 U of NEB Exonuclease III (NEB, #M0206S), and incubated for 5 min at 37 °C to remove biotin from unligated ends. Crosslinks were reversed by adding 31.25 μ L of 20 mg/mL Proteinase K (NEB, P8107) and 25 μ L of 10% SDS (65 °C, overnight).

The chromatin was phenol/chloroform purified and separated on a 1.5% agarose gel. Fragments between 250-400 bp in size were extracted and purified using the NucleoSpin Gel and PCR Clean-up kit (Macherey & Nagel) and resuspended in 56 μ L of elution buffer. For end repair, 15 μ L of end-repair mix was added (7 μ L of 10X T4 DNA Ligase Reaction Buffer (NEB #B0202S), 1 μ L of 25 mM dNTP mix, 2.5 μ L of 10 U/ μ L NEB T4 PNK (NEB, M0201), 2.5 μ L of 3 U/ μ L NEB T4 DNA polymerase I (NEB, M0203), 0.5 μ L of 5 U/ μ L NEB DNA polymerase I, Large (Klenow) Fragment (NEB, M0210), 1.5 μ L of H₂O) and incubated for 30 min at 20 °C and then for 20 min at 75 °C after addition of 0.5 M EDTA, pH 8 to a final concentration of 10 mM.

To isolate biotin-labelled ligation junctions, 50 μ L of 10 mg/mL DynabeadsTM MyOneTM Streptavidin C1 beads (Invitrogen, #65001) were washed with 400 μ L of 1X Tween washing buffer (TWB; 5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween 20), collected with a magnet, resuspended in 400 μ L of 2X binding buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 M NaCl) and added to 330 μ L of TLE (10 mM Tris-HCl (pH 7.5), 0.1 mM EDTA). 70 μ L of the sample were added to the beads and incubated at room temperature for 15 min with slow rotation. The beads were washed once with 400 μ L of 1X TLE and finally resuspended in 41 μ L of 1X TLE.

Sequencing library preparation

For polyadenylation, 5 μ L of 10X NEBuffer 2.1, 1 μ L of 10 mM dATP and 3 μ L of 5 U/ μ L of Klenow fragment (3' \rightarrow 5' exo-) (NEB, M0212) and incubated for 30 min at 37 °C followed by deactivation for 20 min at 65 °C. Beads were reclaimed with a magnet, washed once with 400 μ L of 1X Quick ligation buffer (NEB, M2200) and resuspended in 46.5 μ L of 1X Quick ligation buffer (NEB, M2200). 2.5 μ L of DNA Quick ligase (NEB, M2200) and 0.5 μ L of 50 μ M annealed TruSeq adapters were added and incubated for 1 h at room temperature. Beads were separated on a magnet, resuspended in 400 μ L of 1X TWB (5 mM Tris-HCl, 0.5 M EDTA, 1 M NaCl, 0.05% Tween-20) and washed for 5 min at room temperature with rotation. On the magnet, beads were washed twice

with 200 µL of 1X binding buffer and once with 200 µL of 1X NEBuffer 2.1 and resuspended in 20 uL of 1X NEBuffer 2.1. Each library was amplified in 12 separate reactions of 25 µL. Per reaction, 0.85 µL of 25 µM TruSeg PCR primer cocktail (TruSeg PCR primer cocktail F: 5'-AATGATACGGCGACCACCGAG-3'; TruSeq PCR primer cocktail_R: 5'-CAAGCAGAAGACGGCATACGAG-3'), 12.5 µL of 2X Kapa HiFi HotStart Ready Mix (Kapa Biosystems, KR0370) and 10.15 µL of water were added to 1.5 µL of library bound to the beads. Amplification was performed as follows: 3 min at 95 °C, 8 cycles of 20 s at 98 °C, 30 s at 63 °C and 30 s at 72 °C, 1 cycle of 1 min at 72 °C, hold at 4 °C. The PCR reactions were pooled and the beads were removed from the supernatant using a magnet. Each library was purified by addition of 1.5 volumes of Agencourt AMPure XP beads (Beckman Coulter), according to the manufacturer's instructions. The sample was eluted off beads using 25 µL of 1x TLE buffer, transferred to a fresh tube and the concentration was determined using Qubit (Qubit dsDNA HS Assay Kit, Thermo Fisher) and qPCR (KAPA SYBR FAST qPCR Master Mix, Kapa Biosystems), according to the manufacturer's instructions. Library size distributions were determined on a 5% polyacrylamide gel. The samples were sequenced on an Illumina NextSeq 1000 or NextSeq 2000 on 60 base pair paired end mode.

Chromosome conformation capture data processing and visualization

For visualization and downstream analysis, Micro-C datasets were processed according to the distiller pipeline (https://github.com/mirnylab/distiller-nf). Briefly, reads were mapped to our *T. brucei* Lister 427 version 12 (Tb427v12) genome assembly, working version. Mapping was done with bwa mem with flags -SP. Alignments were parsed and pairs were classified using the pairtools package¹⁰⁰ to generate 4DN-compliant pairs files. Pairs having matching alignment strands and coordinates with a possible 1 bp offset were considered PCR and/ or optical duplicates and thus removed. Pairs classified as uniquely mapped or rescued chimeras with high mapping quality scores on both sides (MAPQ > 30) were corrected for ploidy as described below and then aggregated into contact matrices in the cooler format (cool files) using the cooler package¹⁰¹ at 1 kb and into multiresolution cooler files (mcool files; resolutions: 1 kb, 2 kb, 5 kb, 10 kb, 50 kb, 100 kb).

Given the highly repetitive nature of the *T. brucei* genome, Micro-C datasets were in parallel processed with the mHi-C pipeline¹⁰² to avoid underestimation of the biological signal from genomic repeats. The two ends of each paired read were mapped using bwa (v0.7.17¹⁰³) with parameters set according to the 4DN guidelines. Re-ligation events, too short interactions and PCR duplicates were excluded. The genome was binned at 5 kb and 10 kb; uniquely mapping reads were binned as well, and multi-mapping reads were probabilistically allocated to a bin if the posterior probability of belonging to the bin was at least 0.6. Both uniquely and multi-mapping interactions were aggregated into contact matrices. Ploidy correction of the contact matrix was implemented as described below. All contact matrices were normalized using the iterative correction procedure⁵⁶ and converted to cool files, then extended to multi-resolution cool files (mcool files) and visualized using HiGlass¹⁰⁴.

Ploidy normalization and conversion to entire chromosomes

To analyze sequencing data from the *T. brucei* chromosome core regions, where the two homologous chromosomes are very similar and the aligner cannot confidently assign a read to one allele or the other, we collapsed the two alleles into a single contig in the reference genome. In contrast, the subtelomeric regions of each homologous chromosome were kept separate. This resulted in the mapper assigning reads from two physical chromosomal regions to one core contig, leading to an overestimation of interaction frequencies in the core regions and a discrepancy between the true and measured interaction frequencies. To address this, the overestimation must be corrected

after alignment. Because similar issues arise when ploidy varies between chromosomes, we refer to this process as "ploidy correction".

A ploidy correction script based on Schmidt et al.⁵³ was implemented that 1) duplicates each interaction occurring in a pairs file and 2) splits the interactions coming from the homozygous cores into interactions coming from computationally defined coreA and coreB. Consequently, a computationally phased genome is defined. This way to correct for ploidy assumes that interactions occur preferentially within the same haplotype. Analogous scripts to convert interaction matrix and annotation files were implemented as well. After correcting for ploidy, the resulting files were bioinformatically processed to place interactions from connected contigs consecutively, so that a group of contigs that constitute a chromosome would be considered as such in all the following analyses of Micro-C data. As mentioned above, the working version of the Tb427v12 genome assembly has only one core region per chromosome (corresponding to core A). As a result, the analyses done on the bioinformatically-generated diploid genome (diploid coreA genome) consider the annotated features of core A twice, while features of core B are not taken in consideration in downstream analyses.

Reproducibility of Micro-C libraries

We used HiCRep¹⁰⁵ to do distance-corrected correlations of the various replicates. Correlation is calculated as follows: first, interaction maps are stratified by genomic distances and the correlation coefficients are calculated for each distance separately; afterwards, the reproducibility is determined by a stratum-adjusted correlation coefficient statistic (SCC) by aggregating stratum-specific correlation coefficients using a weighted average. As input files, we used genome-wide matrix files (bin1 bin2 value) at 100 kb resolution. The results are displayed as correlation heat map in Supplementary Fig. 5.

Calling of enriched interactions

Enriched interactions outside the diagonal signal were called using Mustache⁵⁴. At 2 and 5 kb resolution, such enriched contacts were called setting the *p*-value threshold parameter of 0.02 and the sparsity threshold parameter to 0.85. The program outputs the pairs of genomic coordinates representing called contacts and the number of contacts detected was derived from the number of entries in each output file.

Pairwise TSSs enrichment

For each pair of TSSs, all normalized Micro-C matrix entries at the intersection of the pair were averaged. As control region, both annotations were shifted 20 kb downstream to compute a second average of entries as before. Enrichment was calculated by dividing each sample average by its control average. Each pair of TSSs was considered enriched if its average value was at least 1.25-fold the respective control average.

Compartment analysis

Chromosomal compartment profiles were obtained using the FAN-C package⁸ (fanc compartments function) through an eigenvector decomposition procedure on observed-over-expected cis contact maps (Micro-C, matrix obtained according to the distiller pipeline) at 10 kb resolution. Since in our Micro-C data the first eigenvector reflected other features of the heat map such as regions of no mappability (Supplementary Fig. 6), we calculated also the second eigenvector for compartment calling. Eigenvectors were oriented using H2A.Z MNase-ChIP-seq data⁵⁵ enrichment.

Insulation score and domain callers

Insulation score was obtained using the FAN-C package $^{\rm 8}$ (fanc insulation function) and calculated at 2.5, 5 and 10 kb resolution on a Micro-C

matrix obtained according to the distiller pipeline. TAD-like structures were called using the hicFindTADs of the HiCExplorer package¹⁰⁶ at 250 bp resolution and exported in a bed file.

Mappability calculation

Single end reads (read 1) from biological replicate 1, technical replicates 1-3 were aligned against the genome with bwa (v0.7.16¹⁰³) and the mappability calculated as ratio between alignments with mapq \geq 30 and alignments with mapq \geq 0 using bamCompare (deeptools v3.5¹⁰⁷).

Detection of no visibility region

Micro-C data were IC normalized to eliminate factorizable biases. This procedure gives no visibility to regions that do not contain mapped reads and that have high noise (i.e., regions with the fewest number of contacts). The no visibility regions were annotated by extracting all balances with cooler¹⁰¹ balance function and extract bins with no values assigned.

Aggregate Analyses

The Aggregate Peak Analysis (APA) implemented in the GENOVA package¹⁰⁸ averages the signal surrounding a two-dimensional set of locations. The APA of detected enriched interactions was carried out using as input i) contact-objects loaded from the Micro-C and Hi-C mcool files; and ii) the pairs of coordinates of Mustache-called enriched interactions. When providing a list of multiple contact-objects, the GENOVA package also performs a differential analysis. Moreover, to distinguish between intra-chromosomal TSS enriched interactions and inter-chromosomal TSS enriched interactions and inter-chromosomal TSS enriched interactions and inter-chromosomal TSS enriched interactions for regions the pairwise combination of the coordinates of annotated TSSs where i) both elements of each pair belong to the same chromosome; or ii) both elements of each pair belong to different chromosomes. As control, the same coordinates were shifted of 20 kb downstream.

The Cross Spatial Chromatin Analysis (CSCAn) makes an averaged contact analysis for crosswise pairs between elements of a list of genomic location. The CSCAn was carried on lists of single and divergent TSSs and single and convergent TTSs. The results of all aggregate analyses were plotted using the visualise function from the same package.

Aggregated PTU matrices were generated by flipping the interactions of all relevant regions (intra- or inter-chromosomal PTUs) to the forward strand and scaling the regions to the same size. Each region is also extended by 10% of its size in both directions. Aggregated interactions were balanced via cooler with the '--ignore-diags' parameter set to zero.

ChIP-seq data analysis

H2A.Z MNase-ChIP-seq data55, SCC1 ChIP-seq data48 and RNAPII ChIPseq data⁶⁶ were aligned to our Tb427v12 genome assembly, working version, using bwa (v0.7.16¹⁰³). The SCC1 ChIP sample and RNAPII ChIP sample were normalized to their respective inputs (ratio of ChIP to input reads). Aligned reads were ploidy normalized and contigs belonging to the same chromosome merged. After SAM to BAM conversion, the aligned reads were sorted, indexed and duplicates were removed using SAMtools version 1.897. H2A.Z peaks were called using MACS2⁹⁸ with -q 0.01, --maxgap 500, --extsize 300, and broad settings. Peaks in centromeric regions were manually removed. SCC1 peaks were called with standard (narrow) settings and --extsize 200. Metaplots along PTUs and around loci of interests (single and divergent TSSs, single and convergent TTSs) were generated using deeptools v3.5¹⁰⁷ computeMatrix and plotProfile functions. A custom script was implemented to overlap called enriched interactions with annotated features.

Virtual 4C analysis

The virtual 4C function implemented in the cooltools package^{100,109} was used to extract all the interaction frequencies of a given viewpoint with the rest of the genome from the Micro-C mcool files. All virtual 4C analysis used as input mHiC matrices (retaining multi-mapping reads) at 5 kb resolutions (viewpoints: tRNA loci and TSSs) or at 10 kb (centromeric and rDNA loci). The tab-separated output files were plotted using the PyGenomeTracks program¹¹⁰ either against the entire genome or against a subset group of genomic elements (e.g., one tRNA vs all tRNA loci; one rDNA vs all rDNA loci; one centromere vs all centromeric regions).

Statistical peak analysis

Virtual 4 C enrichment *p*-values were computed using a two-sample Kolmogorov–Smirnov test (KS-test). For each genomic locus, a local and a global sample were collected, representing the local and global distributions. The local sample consists of all virtual 4C values within a window of the specific locus, while the global sample contains all other values. The window size was chosen according to the viewpoint size (100 or 500 kb for centromeres, 10 kb for rRNA regions, 5 or 10 kb for tRNA regions). The chromosome of the viewpoint was excluded from this analysis. The null hypothesis for the KS-test was that the local distribution is larger or equal to the global distribution. All values in the local sample were divided by 2 before applying the KS-test to check for a \geq 2-fold enrichment.

Fluorescence In Situ Hybridization (FISH)

For DNA FISH experiments, Digoxigenin- and Alexa Fluor 594-labelled DNA probes were generated by PCR using standard conditions with OneTaq polymerase (New England Biolabs), with a 1:2 ratio of ChromaTideTM Alexa FluorTM 594-5-dUTP (Invitrogen) or Digoxigenin-11-dUTP (Roche) and dNTPs were used in the reaction. CIR147 repeats^{83,84} were amplified from *T. brucei* Lister 427 genomic DNA (Supplementary Table 4). A smear of products with various sizes was generated but only fragments lower than 500 bp were gel extracted and purified with NucleoSpin Gel and PCR Clean-up kit (Macherey & Nagel). To generate Spliced Leader (SL) probes, we divided the SL region into 5 fragments of around 200 bp, then amplified from *T. brucei* Lister 427 genomic DNA (Supplementary Table 4). Probes were gel extracted and purified with NucleoSpin Gel and PCR Clean-up kit (Macherey & Nagel). After gel purification, probes were further purified by ethanol precipitation and resuspended in water.

For each assay, cells were fixed in 2% formaldehyde in media for 10 min at room temperature, washed three times with PBS and finally resuspended in PBS. The cells were attached to poly-llysine-treated coverslips, then permeabilized with 0.5% Triton X-100 in PBS for 15 min at room temperature, washed three times with PBS. Afterward, cells were incubated in 0.1 N HCl for 5 min, washed twice in PBS at room temperature, and then treated with 1 mg/mL RNAse A (Invitrogen) in PBS for 60 min at 37°C. Lastly, cells were washed with PBS three times and incubated overnight in 2X saline-sodium citrate (SSC)/50% formamide/50 mM sodium phosphate at room temperature. Cells were then blocked with hybridization buffer (2X SSC, 5X Denhardt's solution, 50% Formamide, 50 mM Sodium phosphate buffer, 1 mM EDTA, 100 µg/ml salmon sperm DNA, 10% dextran sulfate, pH 7.5-8) for 60 min at 37°C. After the hybridization buffer was removed from the sample, 10 μ L of hybridization solution containing the 1 ng/ μ L probe were added and the coverslip sealed with fixogum (Marabu). Samples were denatured on a heat block at 75 °C for 1 min, followed by overnight incubation at 37 °C. After hybridization, coverslips were washed with three 10-min washes in 4X SSC, 2X SSC and 1X SSC at 50 °C. Samples were then washed three times for 10 min in PBS with 0.05% Tween. Alexa Fluor 594-labelled samples were then mounted in Vectashield with DAPI. Digoxigenin-labelled samples were incubated with a mouse antidigoxigenin antibody (Abcam; clone 21H8) diluted 1:10,000 in 1% BSA in PBS for 2 h at room temperature. After washing three times for 10 min in PBS with 0.05% Tween, digoxigenin-labelled samples were incubated for 1 h with a donkey anti-mouse Alexa Fluor 488 antibody (Invitrogen) diluted to 1:500 in 1% BSA. Samples were washed in PBS with 0.05% Tween and mounted in Vectashield with DAPI.

For imaging, a wide-field fluorescence Leica DMI8 microscope with Lumencor Spectra X light source and a HC PL APO $100 \times /1.47$ OIL objective was used. Images were captured with a Leica DFC9000 GT sCMOS camera. The images were acquired as *z* stacks (0.15 µm) and further deconvolved using the Deconvolf software¹¹¹.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

High-throughput sequencing Micro-C data and ultra-long ONT reads generated in this study have been deposited in the European Nucleotide Archive under primary accession number PRJEB76933 [https:// www.ebi.ac.uk/ena/browser/view/PRJEB76933]. The resulting genome assembly is available at https://zenodo.org/doi/10.5281/zenodo. 12683395. Previously published ChIP-seq and Hi-C data that were used in this study are publicly available at the European Nucleotide Archive or through GEO Series under accession numbers GSE98061 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE98061]

(H2A.Z MNase-ChIP-seq⁵⁵), GSE100896 [http://www.ncbi.nlm.nih.gov/ geo/query/acc.cgi?acc=GSE100896] (SCC1 ChIP-seq⁴⁸, samples GSM3357444 and GSM3357445), GSE150253 [https://www.ncbi.nlm. nih.gov/geo/query/acc.cgi?acc=GSE150253] (RNAPII ChIP-seq⁶⁶, samples GSM5381488 and GSM5381492), PRJEB35632 [https://www.ebi.ac. uk/ena/data/view/PRJEB35632] and GSE100896 [http://www.ncbi.nlm. nih.gov/geo/query/acc.cgi?acc=GSE100896] (Hi-C datasets^{30,48}, merged in this study).

Code availability

The code for the generation and quality control of the genome assembly has been deposited at Zenodo https://zenodo.org/doi/10. 5281/zenodo.12683395. The code for the analysis of Micro-C and Hi-C data is available at https://zenodo.org/doi/10.5281/zenodo.12683439.

References

- 1. Misteli, T. Beyond the Sequence: Cellular Organization of Genome Function. *Cell* **128**, 787–800 (2007).
- Fraser, P. & Bickmore, W. Nuclear organization of the genome and the potential for gene regulation. *Nature* 447, 413–417 (2007).
- De Wit, E. & De Laat, W. A decade of 3C technologies: insights into nuclear organization. Genes Dev. 26, 11–24 (2012).
- Cavalli, G. & Misteli, T. Functional implications of genome topology. Nat. Struct. Mol. Biol. 20, 290–299 (2013).
- Beagan, J. A. & Phillips-Cremins, J. E. On the existence and functionality of topologically associating domains. *Nat. Genet.* 52, 8–16 (2020).
- 6. Lieberman-Aiden, E. et al. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289–293 (2009).
- Lajoie, B. R., Dekker, J. & Kaplan, N. The Hitchhiker's guide to Hi-C analysis: Practical guidelines. *Methods* 72, 65–75 (2015).
- Kruse, K., Hug, C. B. & Vaquerizas, J. M. FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data. *Genome Biol.* 21, 303 (2020).

- Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380 (2012).
- 10. Davidson, I. F. et al. DNA loop extrusion by human cohesin. *Science* **366**, 1338–1345 (2019).
- 11. Rao, S. S. P. et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).
- 12. Hanssen, L. L. P. et al. Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nat. Cell Biol.* **19**, 952–961 (2017).
- Harmston, N. et al. Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nat. Commun.* 8, 441 (2017).
- 14. Jin, F. et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290–294 (2013).
- Mifsud, B. et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* 47, 598–606 (2015).
- Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* 20, 437–455 (2019).
- 17. Uyehara, C. M. & Apostolou, E. 3D enhancer-promoter interactions and multi-connected hubs: Organizational principles and functional roles. *Cell Rep.* **42**, 112068 (2023).
- Wansink, D. et al. Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. J. Cell Biol. 122, 283–293 (1993).
- Šmigová, J., Juda, P., Cmarko, D. & Raška, I. Fine structure of the 'PcG body' in human U-2 OS cells established by correlative lightelectron microscopy. *Nucleus* 2, 219–228 (2011).
- 20. Quinodoz, S. A. et al. Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**, 744–757.e24 (2018).
- Boisvert, F.-M., Van Koningsbruggen, S., Navascués, J. & Lamond, A. I. The multifunctional nucleolus. *Nat. Rev. Mol. Cell Biol.* 8, 574–585 (2007).
- Thompson, M., Haeusler, R. A., Good, P. D. & Engelke, D. R. Nucleolar Clustering of Dispersed tRNA Genes. Science 302, 1399–1401 (2003).
- 23. Raab, J. R. et al. Human tRNA genes function as chromatin insulators: Human tRNA genes function as chromatin insulators. *EMBO J.* **31**, 330–350 (2012).
- 24. Van Bortle, K., Phanstiel, D. H. & Snyder, M. P. Topological organization and dynamic regulation of human tRNA genes during macrophage differentiation. *Genome Biol.* **18**, 180 (2017).
- Jackson, D. A., Iborra, F. J., Manders, E. M. M. & Cook, P. R. Numbers and Organization of RNA Polymerases, Nascent Transcripts, and Transcription Units in HeLa Nuclei. *Mol. Biol. Cell* 9, 1523–1536 (1998).
- Osborne, C. S. et al. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* 36, 1065–1071 (2004).
- Cho, W.-K. et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* 361, 412–415 (2018).
- 28. Hilbert, L. et al. Transcription organizes euchromatin via microphase separation. *Nat. Commun.* **12**, 1360 (2021).
- Mora, A., Huang, X., Jauhari, S., Jiang, Q. & Li, X. Chromatin Hubs: A biological and computational outlook. *Comput. Struct. Biotechnol. J.* 20, 3796–3813 (2022).
- Faria, J. et al. Spatial integration of transcription and splicing in a dedicated compartment sustains monogenic antigen expression in African trypanosomes. *Nat. Microbiol.* 6, 289–300 (2021).
- 31. Budzak, J., Jones, R., Tschudi, C., Kolev, N. G. & Rudenko, G. An assembly of nuclear bodies associates with the active VSG

expression site in African trypanosomes. *Nat. Commun.* **13**, 101 (2022).

- Barcons-Simon, A., Carrington, M. & Siegel, T. N. Decoding the impact of nuclear organization on antigenic variation in parasites. *Nat. Microbiol.* 8, 1408–1418 (2023).
- Michaeli, S. Trans-Splicing in Trypanosomes: Machinery and Its Impact on the Parasite Transcriptome. *Future Microbiol* 6, 459–474 (2011).
- 34. Bhat, P. et al. Genome organization around nuclear speckles drives mRNA splicing efficiency. *Nature* **629**, 1165–1173 (2024).
- 35. Krietenstein, N. et al. Ultrastructural Details of Mammalian Chromosome Architecture. *Mol. Cell* **78**, 554–565.e7 (2020).
- Hsieh, T.-H. S. et al. Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol. Cell* 78, 539–553.e8 (2020).
- Szabo, Q. et al. TADs are 3D structural units of higher-order chromosome organization in *Drosophila*. Sci. Adv. 4, eaar8082 (2018).
- Hsieh, T.-H. S. et al. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. Cell 162, 108–119 (2015).
- Hsieh, T.-H. S., Fudenberg, G., Goloborodko, A. & Rando, O. J. Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. *Nat. Methods* 13, 1009–1011 (2016).
- Mizuguchi, T. et al. Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. *Nature* 516, 432–435 (2014).
- 41. Nagano, T. et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).
- 42. Flyamer, I. M. et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544**, 110–114 (2017).
- 43. Nagano, T. et al. Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).
- Friman, E. T., Flyamer, I. M., Marenduzzo, D., Boyle, S. & Bickmore, W. A. Ultra-long-range interactions between active regulatory elements. *Genome Res.* 33, 1269–1283 (2023).
- 45. Hamley, J. C., Li, H., Denny, N., Downes, D. & Davies, J. O. J. Determining chromatin architecture with Micro Capture-C. *Nat. Protoc.* **18**, 1687–1711 (2023).
- Goel, V. Y., Huseyin, M. K. & Hansen, A. S. Region Capture Micro-C reveals coalescence of enhancers and promoters into nested microcompartments. *Nat. Genet.* 55, 1048–1056 (2023).
- 47. Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E. & Wiehe, T. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc. Natl Acad. Sci.* **109**, 17507–17512 (2012).
- Müller, L. S. M. et al. Genome organization and DNA accessibility control antigenic variation in trypanosomes. *Nature* 563, 121–125 (2018).
- 49. Clayton, C. E. Life without transcriptional control? From fly to man and back again. *EMBO J.* **21**, 1881–1888 (2002).
- 50. Luzak, V. et al. SLAM-seq reveals independent contributions of RNA processing and stability to gene expression in African trypanosomes. *Nucleic Acids Res.* https://doi.org/10.1093/nar/ gkae1203 (2024).
- 51. Cosentino, R. O., Brink, B. G. & Siegel, T. N. Allele-specific assembly of a eukaryotic genome corrects apparent frameshifts and reveals a lack of nonsense-mediated mRNA decay. *NAR Genomics Bioinforma*. **3**, lqab082 (2021).
- 52. Krasiļņikova, M. et al. Nanopore sequencing reveals that DNA replication compartmentalisation dictates genome stability and instability in *Trypanosoma brucei*. Preprint at https://doi.org/10. 1101/2024.05.07.592375 (2024).
- Schmidt, M. R., Barcons-Simon, A., Rabuffo, C. & Siegel, T. N. Smoother: on-the-fly processing of interactome data using prefix sums. *Nucleic Acids Res.* 52, e23–e23 (2024).

Article

- Roayaei Ardakany, A., Gezer, H. T., Lonardi, S. & Ay, F. Mustache: multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation. *Genome Biol.* 21, 256 (2020).
- Wedel, C., Förstner, K. U., Derr, R. & Siegel, T. N. GT-rich promoters can drive RNA pol II transcription and deposition of H2A.Z in African trypanosomes. *EMBO J.* 36, 2581–2594 (2017).
- 56. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
- Díaz-Viraqué, F., Chiribao, M. L., Libisch, M. G. & Robello, C. Genome-wide chromatin interaction map for Trypanosoma cruzi. *Nat. Microbiol.* 8, 2103–2114 (2023).
- Bellini, N. K., De Lima, P. L. C., Pires, D. D. S. & Da Cunha, J. P. C. Hidden origami in *Trypanosoma cruzi* nuclei highlights its nonrandom 3D genomic organization. Preprint at https://doi.org/10. 1101/2024.07.01.601582 (2024).
- 59. Banigan, E. J. et al. Transcription shapes 3D chromatin organization by interacting with loop extrusion. *Proc. Natl Acad. Sci.* **120**, e2210480120 (2023).
- Gluenz, E., Sharma, R., Carrington, M. & Gull, K. Functional characterization of cohesin subunit SCC1 in *Trypanosoma brucei* and dissection of mutant phenotypes in two life cycle stages. *Mol. Microbiol.* 69, 666–680 (2008).
- Landeira, D., Bart, J.-M., Van Tyne, D. & Navarro, M. Cohesin regulates VSG monoallelic expression in trypanosomes. *J. Cell Biol.* 186, 243–254 (2009).
- 62. Siegel, T. N. et al. Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes Dev.* **23**, 1063–1076 (2009).
- 63. Wright, J. R., Siegel, T. N. & Cross, G. A. M. Histone H3 trimethylated at lysine 4 is enriched at probable transcription start sites in Trypanosoma brucei. *Mol. Biochem. Parasitol.* **172**, 141–144 (2010).
- Maree, J. P. & Patterton, H.-G. The epigenome of Trypanosoma brucei: A regulatory interface to an unconventional transcriptional machine. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1839**, 743–750 (2014).
- Kraus, A. J. et al. Distinct roles for H4 and H2A.Z acetylation in RNA transcription in African trypanosomes. *Nat. Commun.* 11, 1498 (2020).
- 66. Staneva, D. P. et al. A systematic analysis of *Trypanosoma brucei* chromatin factors identifies novel protein interaction networks associated with sites of transcription initiation and termination. *Genome Res.* **31**, 2138–2154 (2021).
- 67. Müller-Hübner, L. The role of nuclear architecture in the context of antigenic variation in Trypanosoma brucei. 23534 KB (Universität Würzburg, 2020). https://doi.org/10.25972/OPUS-18707.
- 68. Cheutin, T. et al. Three-dimensional organization of active rRNA genes within the nucleolus. *J. Cell Sci.* **115**, 3297–3307 (2002).
- 69. Moss, T., Langlois, F., Gagnon-Kugler, T. & Stefanovsky, V. A housekeeper with power of attorney: the rRNA genes in ribosome biogenesis. *Cell. Mol. Life Sci.* **64**, 29–49 (2007).
- Sanij, E. et al. UBF levels determine the number of active ribosomal RNA genes in mammals. J. Cell Biol. 183, 1259–1274 (2008).
- Gupta, S. & Santoro, R. Regulation and Roles of the Nucleolus in Embryonic Stem Cells: From Ribosome Biogenesis to Genome Organization. Stem Cell Rep. 15, 1206–1219 (2020).
- Alsford, S., Kawahara, T., Glover, L. & Horn, D. Tagging a T. brucei RRNA locus improves stable transfection efficiency and circumvents inducible expression position effects. *Mol. Biochem. Parasitol.* **144**, 142–148 (2005).
- 73. Ridewood, S. et al. The role of genomic location and flanking 3'UTR in the generation of functional levels of variant surface glycoprotein in *Trypanosoma brucei*. *Mol. Microbiol.* **106**, 614–634 (2017).
- Mottram, J. C., Bell, S. D., Nelson, R. G. & Barry, J. D. tRNAs of Trypanosoma brucei. Unusual gene organization and mitochondrial importation. *J. Biol. Chem.* 266, 18313–18317 (1991).

- International Human Genome Sequencing Consortium. et al. Initial sequencing and analysis of the human genome. *Nature* 409, 860–921 (2001).
- 76. Berriman, M. et al. The Genome of the African Trypanosome. **309**, (2005).
- 77. Pan, T. Modifications and functional genomics of human transfer RNA. *Cell Res.* **28**, 395–404 (2018).
- Cordingley, J. S. Nucleotide sequence of the 5S ribosomal RNA gene repeat of Trypanosoma brucei. *Mol. Biochem. Parasitol.* 17, 321–330 (1985).
- Fantoni, A. RNA Polymerase III-Mediated Transcription of the Trypanosome U2 Small Nuclear RNA Gene Is Controlled by both Intragenic and Extragenic Regulatory Elements. *MOL CELL BIOL* 14, 2021–8 (1994).
- Nakaar, V., Hong, D., Ullu, E. & Tschudil, C. Upstream tRNA Genes Are Essential for Expression of Small Nuclear and Cytoplasmic RNA Genes in Trypanosomes. *MOL CELL BIOL* 14, 6736–6742 (1994).
- 81. Padilla-Mejía, N. E. et al. Gene organization and sequence analyses of transfer RNA genes in Trypanosomatid parasites. *BMC Genomics* **10**, 232 (2009).
- 82. Akiyoshi, B. & Gull, K. Discovery of Unconventional Kinetochores in Kinetoplastids. *Cell* **156**, 1247–1258 (2014).
- Obado, S. O., Taylor, M. C., Wilkinson, S. R., Bromley, E. V. & Kelly, J. M. Functional mapping of a trypanosome centromere by chromosome fragmentation identifies a 16-kb GC-rich transcriptional "strand-switch" domain as a major feature. *Genome Res.* 15, 36–43 (2005).
- 84. Patrick, K. L. et al. Distinct and overlapping roles for two Dicer-like proteins in the RNA interference pathways of the ancient eukaryote *Trypanosoma brucei. Proc. Natl Acad. Sci.* **106**, 17933–17938 (2009).
- Obado, S. O., Bot, C., Nilsson, D., Andersson, B. & Kelly, J. M. Repetitive DNA is associated with centromeric domains in Trypanosoma brucei but not Trypanosoma cruzi. *Genome Biol.* 8, R37 (2007).
- Dekker, J. & Misteli, T. Long-Range Chromatin Interactions. Cold Spring Harb. Perspect. Biol. 7, a019356 (2015).
- Deshpande, A. S. et al. Identifying synergistic high-order 3D chromatin conformations from genome-scale nanopore concatemer sequencing. *Nat. Biotechnol.* **40**, 1488–1499 (2022).
- Cordon-Obras, C. et al. Identification of sequence-specific promoters driving polycistronic transcription initiation by RNA polymerase II in trypanosomes. *Cell Rep.* 38, 110221 (2022).
- Khanna, N., Hu, Y. & Belmont, A. S. HSP70 Transgene Directed Motion to Nuclear Speckles Facilitates Heat Shock Activation. *Curr. Biol.* 24, 1138–1144 (2014).
- 90. Staneva, D. P. et al. The SPARC complex defines RNAPII promoters in Trypanosoma brucei. *eLife* **11**, e83135 (2022).
- 91. Solovei, I. & Mirny, L. Spandrels of the cell nucleus. *Curr. Opin. Cell Biol.* **90**, 102421 (2024).
- Figueiredo, L. M., Janzen, C. J. & Cross, G. A. M. A Histone Methyltransferase Modulates Antigenic Variation in African Trypanosomes. *PLoS Biol.* 6, e161 (2008).
- Zimin, A. V. & Salzberg, S. L. The SAMBA tool uses long reads to improve the contiguity of genome assemblies. *PLOS Comput. Biol.* 18, e1009860 (2022).
- Girasol, M. J. et al. RAD51-mediated R-loop formation acts to repair transcription-associated DNA breaks driving antigenic variation in *Trypanosoma brucei. Proc. Natl Acad. Sci.* **120**, e2309306120 (2023).
- Steinbiss, S. et al. Companion: a web server for annotation and analysis of parasite genomes. *Nucleic Acids Res.* 44, W29–W34 (2016).
- 96. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
- 97. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

Article

- 98. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* **7**, 1728–1740 (2012).
- Benson, G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27, 573–580 (1999).
- Open2C. et al. Pairtools: From sequencing data to chromosome contacts. PLOS Comput. Biol. 20, e1012164 (2024).
- Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* 36, 311–316 (2020).
- Zheng, Y., Ay, F. & Keles, S. Generative modeling of multi-mapping reads with mHi-C advances analysis of Hi-C studies. *eLife* 8, e38070 (2019).
- Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://doi.org/10.48550/ ARXIV.1303.3997 (2013).
- Kerpedjiev, P. et al. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* 19, 125 (2018).
- Yang, T. et al. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.* 27, 1939–1949 (2017).
- Wolff, J. et al. Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* 48, W177–W184 (2020).
- Ramírez, F. et al. deepTools2: a next generation web server for deepsequencing data analysis. *Nucleic Acids Res.* 44, W160–W165 (2016).
- van der Weide, R. H. et al. Hi-C analyses with GENOVA: a case study with cohesin variants. NAR Genomics Bioinforma. 3, lqab040 (2021).
- Open2C et al. Cooltools: enabling high-resolution Hi-C analysis in Python. *bioRxiv* 2022.10.31.514564 https://doi.org/10.1101/2022. 10.31.514564 (2022).
- Lopez-Delisle, L. et al. pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* 37, 422–423 (2021).
- Wernersson, E. et al. Deconwolf enables high-performance deconvolution of widefield fluorescence microscopy images. *Nat. Methods* https://doi.org/10.1038/s41592-024-02294-7 (2024).

Acknowledgements

We thank N. Krietenstein for advice on the Micro-C protocol. We thank M. Oudelaar, I. Solovei and all members of the Siegel, Allshire, Ladurner, Meissner and Boshart laboratories for valuable discussions, T. Straub (Bioinformatics Core Facility, BMC) for providing server space, the Core Unit LAFUGA at the Gene Center Munich for next-generation sequencing and A. Thomae and S. Dietzel (Bioimaging Core Facility, BMC) for providing microscopes. This work was funded by the German Research Foundation (SI 1610 / 2-2 and 213249687—SFB 1064 to T.N.S.), an ERC Starting Grant (3D_Tryps 715466) and an ERC Consolidator Grant (SwitchDecoding 101044320) awarded to T.N.S. P.Y. was supported by an MSCA ITN Cell2Cell (86067) fellowship. The Medical Research Council (UK) supported this research through funding to R.C.A. and K.R.M. (MR/T04702X/1), along with Wellcome Trust funding to R.C.A. (106144 and 225237) and K.R.M. (221717) and to the Discovery Research Platform for Hidden Cell Biology (226791, Bioinformatics core).

Author contributions

The experiments were designed by C.R. and T.N.S. and carried out by C.R. unless otherwise indicated. Genomic DNA extraction for ONT sequencing was carried out by A.B.S., R.C. and S.K. Assembly of the new genome reference and quality controls were performed by M.R.S. with the contribution of R.O.C. Annotation of the new reference genome was performed by M.R.S. and P.T. under the supervision of R.C.A. and K.R.M. Micro-C data were analyzed by C.R. and M.R.S. ChIP-seq data analysis was performed by C.R. and P.T. FISH experiments were performed and analyzed by P.Y. The work was supervised by T.N.S. The manuscript was written by C.R. and T.N.S. and edited by all co-authors. All authors contributed to the data interpretation and development of a model. Figures were generated by M.R.S. and C.R.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-55285-9.

Correspondence and requests for materials should be addressed to T. Nicolai Siegel.

Peer review information *Nature Communications* thanks Peter Myler, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/ licenses/by/4.0/.

© The Author(s) 2024